

# Proceedings of

The 22<sup>nd</sup> Pacific Asia Conference on  
Language, Information and Computation

## PACLIC 22

Dates

November 20-22, 2008

Venue

The University of the Philippines Visayas Cebu College  
Cebu City, Philippines

Hosted By

De La Salle University – Manila

Sponsored by

Philippine Council for Advanced Science and Technology Research and  
Development, Department of Science and Technology

Commission on Higher Education, Philippines

Published by

De La Salle University – Manila (DLSU)

<http://www.dlsu.edu.ph>

(contact information)

Dr. Rachel Roxas

College of Computer Studies

De La Salle University – Manila

2401 Taft Avenue,

Manila, Philippines 1004

[\(roxasr@dlsu.edu.ph\)](mailto:roxasr@dlsu.edu.ph)

All rights reserved. No part of this publication may be reproduced and/or stored in a retrieval system or transmitted, in any form or by any means electronic, mechanical, photocopying, recording or otherwise, without prior permission from the authors.

## **ACKNOWLEDGEMENTS**

PACLIC 22 is financially supported by De La Salle University-Manila, the Philippine Council for Advanced Science and Technology Research and Development - Department of Science and Technology, and the Philippine Commission on Higher Education.

# FOREWORD

It is my great pleasure and honor to present to you the proceedings for the 22<sup>nd</sup> Pacific Asia Conference on Language, Information and Computation which is held for the first time in the Philippines!

I have been eagerly anticipating the hosting of the 22<sup>nd</sup> PACLIC this year 2008 for it also marks the 12<sup>th</sup> year of our involvement in natural language processing (NLP) research in the country. Way back in 1996, Allan Borra and I started a funded project on IsaWika!, a machine translation (MT) system involving the Filipino language. We have come a long way in our NLP research since then. For the past years, we have considered not only MT, but also text summarization, text simplification, natural language generation, information extraction, information retrieval, and language documentation, to name a few.

Secondly, consistent with the vision of PACLIC, we are also happy with the growth of the collaboration between experts on linguistics and computational aspects of languages in the country. This is evidenced by this year's 5<sup>th</sup> National Natural Language Processing Research Symposium, which reflects the zeal and commitment of researchers from all over the country to get involved in this area of research and maximize the potential of bringing these two areas of knowledge together.

And last, but not the least, we have come to realize that our attempts to remain in our comfort zone of textual information had become futile, and we have consequently expanded on input data from pure text to audios (speech) and videos (Filipino sign language).

This growth of our activities and participation is consistently reflected in Pacific Asia researches as evidenced by the papers that are presented in this Conference, and by the topics of our invited and plenary speakers to this conference. We have received a wide variety of papers on linguistics, information, and computation projects of researchers from thirteen countries (namely: Brazil, Canada, China, France, Hong Kong, India, Indonesia, Japan, Philippines, South Korea, Taiwan, United Kingdom, and the United States). Our Program Committee members have done a difficult undertaking in examining the 71 papers submitted to us for evaluation. Through double-blind review, 41% were accepted as regular papers (or 29 papers) and another 20% as posters (or 14 papers), or an overall acceptance rate of 61%. PACLIC Conference proceedings are ISI indexed and is made possible through the coordination of Professor Yasunari Harada through the Waseda University Library Institutional Repository.

Our colleagues from all over the world have graciously accepted our invitation for them to speak at our Conference this year to share with us their expertise in various fields in linguistics, information and computation. We have the privilege of having Mr. Adam Pease as invited speaker, and Professors Danilo Dayag, Trevor Johnston, Chungmin Lee, Haizhou Li, Patrick Saint-Dizier, Sachiko Shudo and Gary Simons as plenary speakers.

I would also like to thank our partners in academe, government, and industry for the support that they have provided for the successful hosting of this Conference.

Gratitude is also expressed to our program committee co-chairs, Professors Hee-Rahk Chae and Tom Lai, and of course, to our local organizing committee headed by Professor Charibeth Cheng, with members Nathalie Rose Lim, and local Cebu partners, Professors Cherry Lyn Sta. Romana and Robert Roxas, and their institutions, the Cebu Institute of Technology and the University of the Philippines Visayas Cebu College (which generously hosted the venue for the conference), and to our staff and students.

I sincerely hope that we will all benefit and be refreshed by the ideas shared in the Conference, and enjoy the beautiful Cebu City, Philippines.

Rachel Edita O. Roxas  
Program Committee Chair, PACLIC 22  
Dean, College of Computer Studies  
De La Salle University-Manila

# CONFERENCE ORGANIZERS

## **Steering Committee of PACLIC**

Jae-Woong Choe, Korea University, Seoul

Yasunari Harada, Waseda University, Tokyo

Chu-Ren Huang, Academia Sinica, Taipei

Kim Teng Lua, Chinese and Oriental Languages Information Processing Society,  
Singapore

Maosong Sun, Tsinghua University, Beijing

Benjamin T'sou, City University of Hong Kong, Hong Kong

## **Program Committee Chair**

Rachel Edita Roxas, De La Salle University-Manila (Chair)

## **Program Committee Co-Chairs**

Hee-Rahk Chae, Hankuk University of Foreign Studies (Co-Chair)

Tom Lai, City University of Hong Kong (Co-Chair)

## **Program Committee**

Sae-Youn Cho (Kangwon National University)

Kazuhiko Fukushima (Kansai Gaidai University)

Takao Gunji (Kobe Shoin Women's University)

Yasunari Harada (Waseda University)

Shu-Kai Hsieh (National Taiwan Normal University)

Kiyoshi Ishikawa (Hosei University)

Dong Hong Ji (Wuhan University, China)

Jong-Bok Kim (Kyung Hee University)

Hiroki Koga (Saga University)

Oi Yee Kwong (City University of Hong Kong)

Minhaeng Lee (Yonsei University)

Qin Lu (The Hong Kong Polytechnic University)

Laurent Prevot (CNRS, Université de Toulouse)  
Byong-Rae Ryu (Chungnam National University)  
Shu-chuan Tseng (Academia Sinica)  
Jie Xu (Huazhong Normal University)  
Eun-Jung Yoo (Seoul National University)  
Ae-Sun Yoon (Pusan National University)  
Quan Zhang (Chinese Academy of Sciences)  
Jun Zhao (Chinese Academy of Sciences)

## **Reviewers**

Winnie Cheng (Hong Kong Polytechnic University)  
Kazuhiko Fukushima (Kansai Gaidai University)  
Takao Gunji (Kobe Shoin Women's University)  
Yasunari Harada (Waseda University)  
Shu-Kai Hsieh (National Taiwan Normal University)  
Chu-Ren Huang (Academia Sinica)  
Kiyoshi Ishikawa (Hosei University)  
Dong Hong Ji (Wuhan University, China)  
Jong-Bok Kim (Kyung Hee University)  
Hiroki Koga (Saga University)  
Oi Yee Kwong (City University of Hong Kong)  
Minhaeng Lee (Yonsei University)  
Qin Lu (The Hong Kong Polytechnic University)  
Thi Minh Huyen Nguyen (Vietnam National University)  
Laurent Prevot (CNRS, Université de Toulouse)  
Byong-Rae Ryu (Chungnam National University)  
Shu-chuan Tseng (Academia Sinica)  
Doreen Wu (Hong Kong Polytechnic University)  
Jie Xu (Huazhong Normal University)  
Eun-Jung Yoo (Seoul National University)  
Quan Zhang (Chinese Academy of Sciences)  
Jun Zhao (Chinese Academy of Sciences)

## **PACLIC22 Local Organizing Committee**

Charibeth Cheng, De La Salle University-Manila (Chair)

Nathalie Rose Lim, De La Salle University-Manila

Robert Roxas, University of the Philippines Visayas Cebu College

Cherry Lyn Sta. Romana, Cebu Institute of Technology

# TABLE OF CONTENTS

## Invited / Plenary Speakers

<b>The Discourse of Print Advertising in the Philippines: Generic Structures and Linguistic Features</b> <i>Danilo Dayag</i>	1
<b>From archive to corpus: transcription and annotation in the creation of signed language corpora</b> <i>Trevor Johnston</i>	16
<b>Scalar Implicatures: Pragmatic Inferences or Grammar?</b> <i>Chungmin Lee</i>	30
<b>NIST 2007 Language Recognition Evaluation: From the Perspective of IIR</b> <i>Haizhou Li, Bin Ma, Kong-Aik Lee, Khe-Chai Sim, Hanwu Sun, Rong Tong, Donglai Zhu, and Changhuai You</i>	46
<b>A Call for Executable Linguistics Research</b> <i>Adam Pease</i>	58
<b>Some Challenges of Advanced Question-Answering: an Experiment with How-to Questions</b> <i>Patrick Saint-Dizier</i>	65
<b>How <i>Even</i> Revises Expectation in a Scalar Model: Analogy with Japanese <i>Mo</i></b> <i>Sachiko Shudo</i>	74
<b>Toward a Global Infrastructure for the Sustainability of Language Resources</b> <i>Gary Simons and Steven Bird</i>	87

## Full Papers

<b>A Rule-based Morpho-semantic Analyzer of the Japanese Verb Phrases of Simple Sentences</b> <i>Yukiko Sasaki Alam</i>	101
<b>Constituent Structure for Filipino: Induction through Probabilistic Approaches</b> <i>Daniel Alcantara and Allan Borra</i>	113
<b>Who's Missing in the Group? Argument Sharing in Core Cosubordinate Construction in Filipino</b> <i>Aquiles Bazar III</i>	123
<b>Towards a Model for the Prediction of Chinese Novel Verbs</b> <i>Paul Chang and Kathleen Ahrens</i>	131
<b>Natural Language Generation of Museum Object Descriptions based on User Model</b> <i>Hsiao Wei Chen, Mary Grace Lim, Patricia Bea Perez, Joanna Patricia Reyes and Nathalie Rose Lim</i>	141
<b>Source of Individuation in Mandarin Chinese, a Classifier Language</b> <i>Pierina Cheung, Peggy Li and David Barner</i>	151



<b>How to Overcome the Domain Barriers in Pattern-Based Machine Translation System</b>	161
<i>Sung-Kwon Choi, Ki-Young Lee, Yoon-Hyung Roh, Oh-Woog Kwon and Young-Gil Kim</i>	
<b>Multi-Engine Approach for Named Entity Recognition in Bengali</b>	169
<i>Asif Ekbal and Sivaji Bandyopadhyay</i>	
<b>Incorporation of WordNet Features to n-gram Features in a Language Modeler</b>	179
<i>Kathleen Go and Solomon See</i>	
<b>An Ontology of Chinese Radicals: Concept Derivation and Knowledge Representation based on the Semantic Symbols of the Four Hoofed-Mammals</b>	189
<i>Chu-Ren Huang, Ya-Jun Yang and Sheng-Yi Chen</i>	
<b>Typology of Comparatives</b>	197
<i>Satomi Ito</i>	
<b>Unsupervised Chinese Verb Metaphor Recognition Based on Selectional Preferences</b>	207
<i>Yuxiang Jia and Shiwen Yu</i>	
<b>Korean Parsing Based on the Applicative Combinatory Categorical Grammar</b>	215
<i>Juyeon Kang and Jean-Pierre Descles</i>	
<b>Stress Processing Sensitivity in Reading Korean and English Words</b>	225
<i>Yongsoon Kang, Seunghyun Baek and Mira Yim</i>	
<b>A Preliminary Study on the Impact of Lexical Concreteness on Word Sense Disambiguation</b>	235
<i>Oi Yee Kwong</i>	
<b>Constructing an Ontology of Coherence Relations: An example of ‘causal relation’</b>	245
<i>Hae-Yun Lee and Sueun Jun</i>	
<b>An Improved Corpus Comparison Approach to Domain Specific Term Recognition</b>	253
<i>Xiaoyue Liu and Chunyu Kit</i>	
<b>Extending an Indonesian Semantic Analysis-based Question Answering System with Linguistic and World Knowledge Axioms</b>	262
<i>Rahmad Mahendra, Septina Dian Larasati and Ruli Manurung</i>	
<b>An Implementation of a Flexible Author-Reviewer Model of Generation using Genetic Algorithms</b>	272
<i>Ruli Manurung, Graeme Ritchie and Henry Thompson</i>	
<b>Semantic Change of the Selected Cebuano Words</b>	282
<i>Rowanne Marie Maxilom</i>	
<b>On Japanese Desiderative Constructions</b>	290
<i>Akira Ohtani and Mark Steedman</i>	
<b>Generating Story Reviews Using Phrases Expressing Emotion</b>	302
<i>Hiroshi Ota and Kazuhide Yamamoto</i>	
<b>Integrating Prosodics into a Language Model for Spoken Language Understanding of Thai</b>	311
<i>Siripong Potisuk</i>	

<b>Using a Word-Space Model to Determine the Relevance of Messages in Anchored Asynchronous Online Discussions</b>	<b>321</b>
<i>Rodolfo Raga, Jennifer Raga, Erick Bonus and Raymund Sison</i>	
<b>Trend-based Document Clustering for Sensitive and Stable Topic Detection</b>	<b>331</b>
<i>Yoshihide Sato, Harumi Kawashima, Hidenori Okuda and Masahiro Oku</i>	
<b>Sentiment Sentence Extraction Using a Hierarchical Directed Acyclic Graph Structure and a Bootstrap Approach</b>	<b>341</b>
<i>Kazutaka Shimada, Daigo Hashimoto and Tsutomu Endo</i>	
<b>An Effective Speech Understanding Method with a Multiple Speech Recognizer based on Output Selection using Edit Distance</b>	<b>350</b>
<i>Kazutaka Shimada, Satomi Horiguchi and Tsutomu Endo</i>	
<b>The Relationship between Semantic Similarity and Subcategorization Frames in English: A Stochastic Test Using ICE-GB and WordNet</b>	<b>358</b>
<i>Sanghoun Song and Jae-Woong Choe</i>	
<b>Automatic Bilingual Lexicon Extraction for a Minority Target Language</b>	<b>368</b>
<i>Eileen Pamela Tiu and Rachel Edita Roxas</i>	

## Posters

<b>Using ‘Low-cost’ Learning Features for Pronoun Resolution</b>	<b>377</b>
<i>Ramon Cuevas and Ivandre Paraboni</i>	
<b>Natural Language Database Interface for the Community Based Monitoring System</b>	<b>384</b>
<i>Krissanne Kaye Garcia, Ma. Angelica Lumain, Jose Antonio Wong, Jhovee Gerard Yap and Charibeth Cheng</i>	
<b>Controlled Korean for Korean-English MT</b>	<b>391</b>
<i>Munpyo Hong and Chang-Hyun Kim</i>	
<b>Statistical Analysis on Large Scale Chinese Short Message Corpus and Automatic Short Message Error Correction</b>	<b>397</b>
<i>Rile Hu, Yuezhong Tang, Chen Li and Xia Wang</i>	
<b>Contrastive Approach towards Text Source Classification based on Top-Bag-of-Word Similarity</b>	<b>404</b>
<i>Chu-Ren Huang and Lung-Hao Lee</i>	
<b>Extracting Troubles from Daily Reports based on Syntactic Pieces</b>	<b>411</b>
<i>Yoshifumi Kakimoto and Kazuhide Yamamoto</i>	
<b>What is Needed the Most in MT-Supported Paper Writing</b>	<b>418</b>
<i>Chang Hyun Kim, Oh-Woog Kwon and Young Kil Kim</i>	
<b>Semantic Structures of Polysemous Psych-adjectives in Korean: A Conceptual Semantics Approach</b>	<b>428</b>
<i>Ilkyu Kim</i>	
<b>Sign language and Computing in a Developing Country: A Research Roadmap for the Next Two Decades in the Philippines</b>	<b>438</b>
<i>Liza Martinez and Ed Peter Cabalfin</i>	
<b>Unsupervised Approach for Dialogue Act Classification</b>	<b>445</b>
<i>Kiyonori Ohtake</i>	

<b>Automatically Extracting Templates from Examples for NLP Tasks</b>	<b>452</b>
<i>Ethel Ong, Bryan Anthony Hong and Vince Andrew Nuñez</i>	
<b>Recognizing Coordinate Structures for Machine Translation of English Patent Documents</b>	<b>460</b>
<i>Yoon-Hyung Roh, Ki-Young Lee, Sung-Kwon Choi, Oh-Woog Kwon and Young-Gil Kim</i>	
<b>A Morphological Analyzer for Filipino Verbs</b>	<b>467</b>
<i>Robert Roxas and Gersam Mula</i>	
<b>Chunking with Max-Margin Markov Networks</b>	<b>474</b>
<i>Tang Buzhou, Wang Xuan and Wang Xiaolong</i>	